Human Adeno-Associated Virus Type 5 Is Only Distantly Related to Other Known Primate Helper-Dependent Parvoviruses

URSULA BANTEL-SCHAAL,* HAJO DELIUS, RAINER SCHMIDT, AND HARALD ZUR HAUSEN

Deutsches Krebsforschungszentrum Heidelberg, Forschungsschwerpunkt Angewandte Tumorvirologie F0400, D-69120 Heidelberg, Germany

We have characterized 95% (4,404 nucleotides) of the genome of adeno-associated virus type 5 (AAV5), including part of the terminal repeats and the terminal resolution site. Our results show that AAV5 is different from all other described AAV serotypes at the nucleotide level and at the amino acid level. The sequence homology to AAV2, AAV3B, AAV4, and AAV6 at the nucleotide level is only between 54 and 56%. The positive strand contains two large open reading frames (ORFs). The left ORF encodes the nonstructural (Rep) proteins, and the right ORF encodes the structural (Cap) proteins. At the amino acid level the identities with the capsid proteins of other AAVs range between 51 and 59%, with a high degree of heterogeneity in regions which are considered to be on the exterior surface of the viral capsid. The overall identity for the nonstructural Rep proteins at the amino acid level is 54.4%. It is lowest at the C-terminal 128 amino acids (10%). There are only two instead of the common three putative Zn fingers in the Rep proteins. The Cap protein data suggest differences in capsid surfaces and raise the possibility of a host range distinct from those of other parvoviruses. This may have important implications for AAV vectors used in gene therapy.

Adeno-associated viruses (AAVs) are small, nonenveloped viruses that encapsidate single-stranded DNA of both polarities in equal amounts. They belong to the *Parvoviridae* family and are distinct from other members of this family by their dependence on helpers for replication (for reviews, see references 7–10, 31, and 37). Six primate AAV serotypes have been reported in the literature (2, 5, 42). They are designated types 1 to 6 (AAV1 to AAV6). With the exception of AAV5, which has been isolated from a penile flat condylomatous lesion (5, 19), all known AAVs were first found as contaminants in laboratory adenovirus stocks (1, 29, 34, 42). Up to now, the DNAs of AAV2, AAV3, AAV4, and AAV6 have been sequenced (17, 36, 41, 42, 51). The sequence identities among the different serotypes are high. The identities within the genomes of AAV2, AAV3, and AAV6 are 82%, and with AAV4 they still range from 75 to 78% (17, 36, 42). For AAV3, two sequences (designated AAV3A and 3B) have been published which have differences in 16 nucleotides (36, 42). In the group of autonomous parvoviruses, the closest relative appears to be the goose parvovirus. At the genomic level and at the level of the capsid proteins, homologies with sequenced AAVs of ca. 54% have been reported (17, 36, 62), and for the nonstructural proteins of AAV3 the identities are ca. 44% (36).

Two large open reading frames (ORFs) have been identified within the AAV genome (7, 8, 17, 36, 37, 41, 42, 51). Experimental data on translation and transcription have mainly been obtained for AAV2, but predictions based on nucleotide sequence analogy could be made for the other AAV serotypes. The left ORF of AAV2 encodes the nonstructural Rep proteins that are transcribed from two separate promoters (p5 and p19, according to their relative map positions). The transcripts from both promoters are translated from spliced and unspliced mRNAs, resulting in four proteins designated Rep78, Rep68, Rep52, and Rep40. The Rep proteins are regulators of AAV transcription; they are involved in multiple steps of AAV replication, and they play a role in the production of single-stranded progeny genomes and virus assembly (7–10, 13–16, 27, 30, 33, 37, 38, 43, 55, 56, 60). In addition, Rep proteins are required for site-specific integration of AAV DNA into the host cell genome (44, 45, 58). Furthermore, they are able to modulate transcription from heterologous promoters (6, 9, 21–23, 25, 26, 28, 32, 37, 61). The degree of sequence conservation for the Rep proteins is high among AAV2, AAV3A, AAV3B, AAV4, and AAV6. The Rep78 proteins from these viruses are reported to be 89 to 93% identical to each other (17, 36, 42). This is thought to mirror their important basic functions in the AAV life cycle (7–10, 17, 37).

The AAV *cap* gene is located in the right half of the AAV genome and codes for the three capsid proteins VP1, VP2, and VP3, with VP3 being the smallest but most abundant; VP1 has the highest molecular weight but is present in a much smaller quantity (as shown for AAV2, AAV3, and AAV5 [7–10, 19]). The respective mRNA is translated from the p40 promoter. As has been shown for AAV2, the Cap proteins differ from each other due to alternative splicing and by the use of an unusual start codon (ACG) for VP2 (7–11, 37). In contrast to the Rep proteins, the reported degree of sequence conservation among the capsid proteins is smaller. This is likely to provide a basis for differences in host range and host cell specificities (17, 36, 37, 42).

AAVs are interesting for several reasons. First of all, they have oncosuppressive properties (3, 22–26, 52, 57; for a review, see reference 40), and they are useful as general transduction vectors for gene therapeutic approaches in human cells (for reviews, see references 31, 37, and 48). Much work has been done with vectors derived from AAV2 (31, 37). However, vectors derived from other AAV serotypes could provide several additional advantages, including the dependence on different cell receptors, resulting in transduction into different cell types, and the resistance to neutralizing antibodies directed against AAV2 (17, 35, 36, 42, 53).

Here we report the partial sequence of AAV5 covering

---

* Corresponding author. Mailing address: Deutsches Krebsforschungszentrum Heidelberg, Forschungsschwerpunkt Angewandte Tumorvirologie F0400, Im Neuenheimer Feld 242, D-69120 Heidelberg, Germany. Phone: 49-6221-424823. Fax: 49-6221-424822. E-mail: u.bantel-schaal@dkfz-heidelberg.de.

```
        10        20        30        40        50        60        70        80        90       100
         |         |         |         |         |         |         |         |         |         |
                   trs
   1 CGCGACAGGG GGGAGAGTGC CACACTCTCA AGCAAGGGGG TTTTGTAAGC AGTGATGTCA TAATGATGTA ATGCTTATTG TCACGCGATA GTTAATGATT
                                                                                            p5
 101 AACAGTCATG TGATGTGTTT TATCCAATAG GAAGAAAGCG CGCGTATGAG TTCTCGCGAG ACTTCCGGGG TATAAAAGAC CGAGTGAACG AGCCCGCCGC
                                                       StartRep78/68
 201 CATTCTTTGC TCTGGACTGC TAGAGGACCC TCGCTGCCAT GGCTACCTTC TATGAAGTCA TTGTTCGCGT CCCATTTGAC GTGGAGGAAC ATCTGCCTGG
 301 AATTTCTGAC AGCTTTGTGG ACTGGGTAAC TGGTCAAATT TGGGAGCTGC CTCCAGAGTC AGATTTAAAT TTGACTCTGG TTGAACAGCC TCAGTTGACG
 401 GTGGCTGATA GAATTCGCCG CGTGTTCCTG TACGAGTGGA ACAAATTTTC CAAGCAGGAG TCCAAATTCT TTGTGCAGTT TGAAAAGGGA TCTGAATATT
 501 TTCATCTGCA CACGCTTGTG GAGACCTCCG GCATCTCTTC CATGGTCCTC GGCCGCTACG TGAGTCAGAT TCGCGCCCAG CTGGTGAAAG TGGTCTTCCA
 601 GGGAATTGAA CCCCAGATCA ACGACTGGGT CGCCATCACC AAGGTAAAGA AGGGCGGAGC CAATAAGGTG GTGGATTCTG GGTATATTCC CGCCTACCTG
                                                               p19
 701 CTGCCGAAGG TCCAACCGGA GCTTCAGTGG GCGTGGACAA ACCTGGACGA GTATAAAATG GCCGCCCTGA ATCTGGAGGA GCGCAAACGG CTCGTCGCGC
                                                                                          StartRep52/40
 801 AGTTTCTGGC AGAATCCTCG CAGCGCTCGC AGGAGGCGGC TTCGCAGCGT GAGTTCTCGG CTGACCCGGT CATCAAAAGC AAGACTTCCC AGAAATACAT
 901 GGCGCTCGTC AACTGGCTCG TGGAGCACGG CATCACTTCC GAGAAGCAGT GGATCCAGGA AAATCAGGAG AGCTACCTCT CCTTCAACTC CACCGGCAAC
1001 TCTCGGAGCC AGATCAAGGC CGCGCTCGAC AACGCGACCA AAATTATGAG TCTGACAAAA AGCGCGGTGG ACTACCTCGT GGGGAGCTCC GTTCCCGAGG
1101 ACATTTCAAA AAACAGAATC TGGCAAATTT TTGAGATGAA TGGCTACGAC CCGGCCTACG CGGGATCCAT CCTCTACGGC TGGTGTCAGC GCTCCTTCAA
1201 CAAGAGGAAC ACCGTCTGGC TCTACGGACC CGCCACGACC GGCAAGACCA ACATCGCGGA GGCCATCGCC CACACTGTGC CCTTTTACGG CTGCGTGAAC
1301 TGGACCAATG AAAACTTTCC CTTTAATGAC TGTGTGGACA AAATGCTCAT TTGGTGGGAG GAGGGAAAGA TGACCAACAA GGTGGTTGAA TCCGCCAAGG
1401 CCATCCTGGG GGGCTCAAAG GTGCGGGTCG ATCAGAAATG TAAATCCTCT GTTCAAATTG ATTCTACCCC TGTCATTGTA ACTTCCAATA CAAACATGTG
1501 TGTGGTGGTG GATGGGAATT CCACGACCTT TGAACACCAG CAGCCGCTGG AGGACCGCAT GTTCAAATTT GAACTGACTA GCGGCTCCC GCCAGATTTT
1601 GGCAAGATTA CTAAGCAGGA AGTCAAGGAC TTTTTTGCTT GGGCAAAGGT CAATCAGGTG CCGGTGACTC ACGAGTTTAA AGTTCCCAGG GAATTGGCGG
                                                                                            p40
1701 GAACTAAAGG GGCGGAGAAA TCTCTAAAAC GCCCACTGGG TGACGTCACC AATACTAGCT ATAAAGTCT GGAGAAGCGG GCCAGGCTCT CATTTGTTCC

        splice --->
1801 CGAGACGCCT CGCAGTTCAG ACGTGACTGT TGATCCCGCT CCTCTGCGAC CGCTCAATTG GAATTCAAGG TATGATTGCA AATGTGACTA TCATGCTCAA
1901 TTTGACAACA TTTCTAACAA ATGTGATGAA TGTGAATATT TGAATCGGGG CAAAAATGGA TGTATCTGTC ACAATGTAAC TCACTGTCAA ATTTGTCATG
                                                                       Stop Rep78/52    <--spl. StartVP1
2001 GGATTCCCCC CTGGGAAAAG GAAAACTTGT CAGATTTTGG GGATTTTGAC GATGCCAATA AAGAACAGTA AATAAAGCGA GTAGTCATGT CTTTTGTTGA
        <--- splice                  Stop Rep68/40
2101 TCACCCTCCA GATTGGTTGG AAGAAGTTGG TGAAGGTCTT CGCGAGTTTT TGGGCCTTGA AGCGGGCCCA CCGAAACCAA AACCCAATCA GCAGCATCAA
2201 GATCAAGCCC GTGGTCTTGT GCTGCCTGGT TATAACTATC TCGGACCCGG AAACGGTCTC GATCGAGGAG AGCCTGTCAA CAGGGCAGAC GAGGTCGCGC
2301 GAGAGCACGA CATCTCGTAC AACGAGCAGC TTGAGGCGGG AGACAACCCC TACCTCAAGT ACAACCACGC GGACGCCGAG TTTCAGGAGA AGCTCGCCGA
                                                                                            StartVP2
2401 CGACACATCC TTCGGGGGAA ACCTCGGAAA GGCAGTCTTT CAGGCCAAGA AAAGGGTTCT CGAACCTTTT GGCCTGGTTG AAGAGGGTGC TAAGACGGCC
2501 CCTACCGGAA AGCGGATAGA CGACCACTTT CCAAAAAGAA AGAAGGCTCG GACCGAAGAG GACTCCAAGC CTTCCACCTC GTCAGACGCC GAAGCTGGAC
                                        StartVP3
2601 CCAGCGGATC CCAGCAGCTG CAAATCCCAG CCCAACCAGC CTCAAGTTTG GGAGCTGATA CAATGTCTGC GGGAGGTGGC GGCCCATTGG GCGACAATAA
2701 CCAAGGTGCC GATGGAGTGG GCAATGCCTC GGGAGATTGG CATTGCGATT CCACGTGGAT GGGGGACAGA GTCGTCACCA AGTCCACCCG AACCTGGGTG
2801 CTGCCCAGCT ACAACAACCA CCAGTACCGA GAGATCAAAA GCGGCTCCGT CGACGGAAGC AACGCCAACG CCTACTTTGG ATACAGCACC CCCTGGGGGT
2901 ACTTTGACTT TAACCGCTTC CACAGCCACT GGAGCCCCCG AGACTGGCAA AGACTCATCA ACAACTACTG GGGCTTCAGA CCCCGGTCCC TCAGAGTCAA
3001 AATCTTCAAC ATTCAAGTCA AAGAGGTCAC GGTGCAGGAC TCCACCACCA CCATCGCCAA CAACCTCACC TCCACCGTCA AGTGTTTAC GGACGACGAC
3101 TACCAGCTGC CCTACGTCGT CGGCAACGGG ACCGAGGGAT GCCTGCCGGC CTTCCCTCCG CAGGTCTTTA CGCTGCCGCA GTACGGTTAC GCGACGCTGA
3201 ACCGCGACAA CACAGAAAAT CCCACCGAGA GGAGCAGCTT CTTCTGCCTA GAGTACTTTC CCAGCAAGAT GCTGAGAACG GGCAACAACT TTGAGTTTAC
3301 CTACAACTTT GAGGAGGTGC CCTTCCACTC CAGCTTCGCT CCCAGTCAGA ACCTCTTCAA GCTGGCCAAC CCGCTGGTGG ACCAGTACTT GTACCGCTTC
3401 GTGAGCACAA ATAACACTGG CGGAGTCCAG TTCAACAAGA ACCTGGCCGG GAGATACGCC AACACCTACA AAAACTGGTT CCCGGGGCCC ATGGGCCGAA
3501 CCCAGGGCTG GAACCTGGGC TCCGGGGTCA ACCGCGCCAG TGTCAGCGCC TTCGCCACGA CCAATAGGAT GGAGCTCGAG GGCGCGAGTT ACCAGGTGCC
3601 CCCGCAGCCG AACGGCATGA CCAACAACCT CCAGGGCAGC AACACCTATG CCCTGGAGAA CACTATGATC TTCAACAGCC AGCCGGCGAA CCCGGGCACC
```
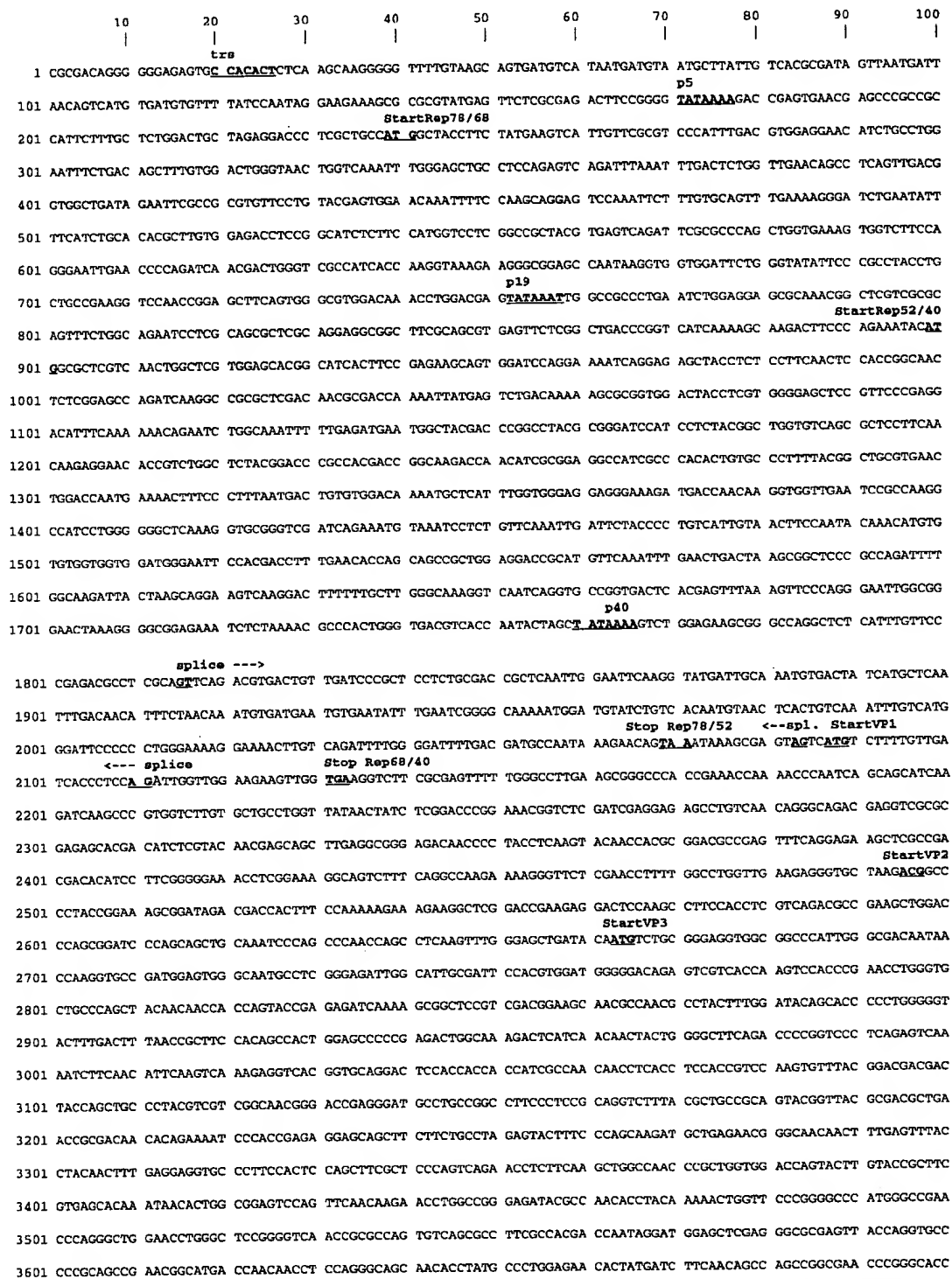
FIG. 1. Partial sequence of AAV5. The 4,404 bases obtained by sequencing of AAV5 are shown. The putative promoters (p5, p19, and p40), polyadenylation signal (polyA), start and stop codons for nonstructural (Rep) and capsid (VP) proteins, splice sites (arrows), and the trs are underlined and marked by boldface letters. The position of the inverted terminal sequence counterpart of the terminal resolution site is also indicated by "trs."

about 95% of the AAV5 genome (4,404 nucleotides [nt]), with the exception of the terminal hairpin structures but including the terminal resolution site and its inverted counterpart (50). The data show that the genomic organization of AAV5 is similar to that of other AAVs but that the interserotype ho-

mology is reduced to values of between 54 and 56% when AAV5 is included into alignments. The differences concern both ORFs. The overall percentage of identical amino acids in the structural proteins is less than 45% and, in contrast to the case with other AAVs, the overall percentage of identical

```
3701 ACCGCCACGT ACCTCGAGGG CAACATGCTC ATCACCAGCG AGAGCGAGAC GCAGCCGGTG AACCGCGTGG CGTACAACGT CGGCGGGCAG ATGGCCACCA

3801 ACAACCAGAG CTCCACCACT GCCCCCGCGA CCGGCACGTA CAACCTCCAG GAAATCGTGC CCGGCAGCGT GTGGATGGAG AGGGACGTGT ACCTCCAAGG

3901 ACCCATCTGG GCCAAGATCC CAGAGACGGG GGCGCACTTT CACCCCTCTC CGGCCATGGG CGGATTCGGA CTCAAACACC CACCGCCCAT GATGCTCATC

4001 AAGAACACGC CTGTGCCCGG AAATATCACC AGCTTCTCGG ACGTGCCCGT CAGCAGCTTC ATCACCCAGT ACAGCACCGG GCAGGTCACC GTGGAGATGG

4101 AGTGGGAGCT CAAGAAGGAA AACTCCAAGA GGTGGAACCC AGAGATCCAG TACACAAACA ACTACAACGA CCCCCAGTTT GTGGACTTTG CCCCGGACAG
                                                                    Stop VPs                                polyA

4201 CACCGGGGAA TACAGAACCA CCAGACCTAT CGGAACCCGA TACCTTACCC GACCCCTTTA ACCCATTCAT GTCGCATACC CTCAATAAAC CGTGTATTCG
                                                                                                  trs

4301 TGTCAGTAAA ATACTGCCTC TTGTGGTCAT TCAATGAATA ACAGCTTACA ACATCTACAA AACCTCCTTG CTTGAGAGTG TGGCACTCTC CCCCCTGTCG

4401 CGCG
```

amino acids encoded by the left-sided ORF is also strongly reduced from 83.4 to 54.4%. Thus, AAV5 is clearly distinct from the other known AAV serotypes.

MATERIALS AND METHODS

Cell culture and virus stocks. HeLa cells were cultured as monolayers in Dulbecco's modified Eagle medium (DMEM; Sigma, Deisenhofen, Germany), supplemented with 5% heat-treated fetal calf serum and glutamine and penicillin-streptomycin at standard concentrations. These cell cultures were used to propagate AAV5 with the helper-adenovirus type 2 (5). Preparations of AAV5 stocks were essentially done as described previously (5), except for the ammonium sulfate precipitation, which was replaced by a centrifugation step at 13,000 rpm for 60 min in the Sorvall SS34 rotor (Sorvall Instruments).

Preparation of viral DNA and cloning of restriction fragments. Viral DNA was isolated from purified virus particles by using alkaline conditions (0.1 N NaOH for 45 min at room temperature). The solution was neutralized, and the DNA was purified with the Geneclean II kit for removing proteins (Bio 101, Inc., La Jolla, Calif.). Restriction fragments (BamHI, EcoRI, SacI, and XhoI) were cloned into bacterial plasmids by standard protocols. pBluescript II(KS+) and pUC18 were used as bacterial vectors for cloning in the bacterial strains HB101 and XL1-Blue.

DNA sequencing. The major part of the sequence determination was done radioactively on plasmid clones of AAV5 by using the dideoxy termination method (46) with general vector primers and primers derived from the 3' part of the newly determined sequences. The terminal sequences and sequences showing deviations between different subclones were determined directly on the viral DNA by cycle sequencing by using the fluorescent dye terminator method (ABI PRISM Big Dye ready reaction terminator cycle sequencing kit) on a model 377 automatic sequencer (Perkin-Elmer/Applied Biosystems) according to the manufacturer's protocol. The partial terminal sequences were confirmed by LION Bioscience, Heidelberg, Germany.

Sequences were aligned with NCBI's MACAW program (Multiple Alignment Construction and Analysis Workbench). Either blocks of similarity or identities of the aligned sequences were shaded.

RESULTS

Nucleotide sequence and genomic organization. The partial AAV5 genome is presented in Fig. 1 and is also available through the EMBL databank.

The major part of the sequence was determined by radioactive sequencing of cloned restriction fragments derived from viral DNA isolated from purified AAV5 particles. Sequence differences in overlapping parts of the subclones were resolved by fluorescent cycle sequencing directly on the single-stranded viral DNA. Viral DNA containing the termini could neither be stably propagated in bacteria nor could the terminal hairpin structures be sequenced by cycle sequencing, since the problem of polymerase stalling in the palindromic structure could not be resolved.

By analogy to other published AAV sequences the sequence presented here (Fig. 1) includes the terminal resolution site (trs [50]) and part of the inverted terminal structures which end at the 5' and 3' ends at symmetrical positions (Fig. 2). As far as the sequence was determined, the 5' and 3' inverted repeat stretches are identical except for a G at position 38, where a T at position 4365 replaced the expected C (Fig. 2A).

A

```
AAV5   nt    1- 47    5' ??----CGCGACAGGGGGGAGAGTGCCACACTCTCAAGCAAGGgGGTTTTGTA --- 3'

AAV5   nt  4404-4356  3' ??--GCGCGCTGTCCCCCCTCTCACGGTGTGAGAGTTCGTTCCtCCAAAACAT --- 5'
```

B

```
AAV2   5'CGCG-CAGAGAGG-GAGTGGCCA-ACTCCATCA --//-- TGATGGAGT-TGGCCACTC-CCTCTCTG-CGCG 3'
AAV3B  5'TGCG-CATAGAGG-GAGTGGCCA-ACTCCATCA --//-- TGATGGAGT-TGGCCACTC-CCTCTATG-CGCA 3'
AAV4   5'CGCG-CATAGAGG-GAGTGGCCA-ACTCCATCA --//-- TGATGGAGT-TGGCCACAT-TAGCTATG-CGCG 3'
AAV6   5'CGCG-CAGAGAGG-GAGTGGCCA-ACTCCATCA --//-- TGATGGAGT-TGCCCACTC-CCTCTATG-CGCG 3'
AAV5   5'CGCGACAGGGGGGAGAGT-GCCACACTC--TCA --//-- TGA--GAGTGTGGC-ACTCTCCCCCCTGTCGCG 3'
```

FIG. 2. Partial sequence of inverted terminal repeats of AAV5. (A) Part of the inverted terminal repeats comprising the first 47 nt that could be deduced at the 5' end (nt 1 to 47) and the last 49 nt of the 3' end (nt 4356 to 4404). The lowercase letters "g" and "t" mark the nucleotide pair at positions 38 and 4365 that does not match the repeat. The putative trs (50) and the respective inverted sequence are shown in boldface letters. Note that the underlined sequences may fold back in AAV5 DNA and form a 7-bp double-stranded loop. (B) Alignment of AAV sequences surrounding the trs and the respective inverted positions in the 5' and 3' ends of the DNA molecules. The trs consensus sequence and the respective inverted sequences are shown in boldface letters. Dashes indicate gaps introduced in the alignment. Nucleotides included in the alignment were as follows: AAV2, 104 to 133 and 4547 to 4576; AAV3B, 103 to 132 and 4590 to 4619; AAV4, 104 to 133 and 4635 to 4664; AAV6, 104 to 133 and 4551 to 4580; AAV5 partial sequence, 1 to 30 and 4373 to 4402.
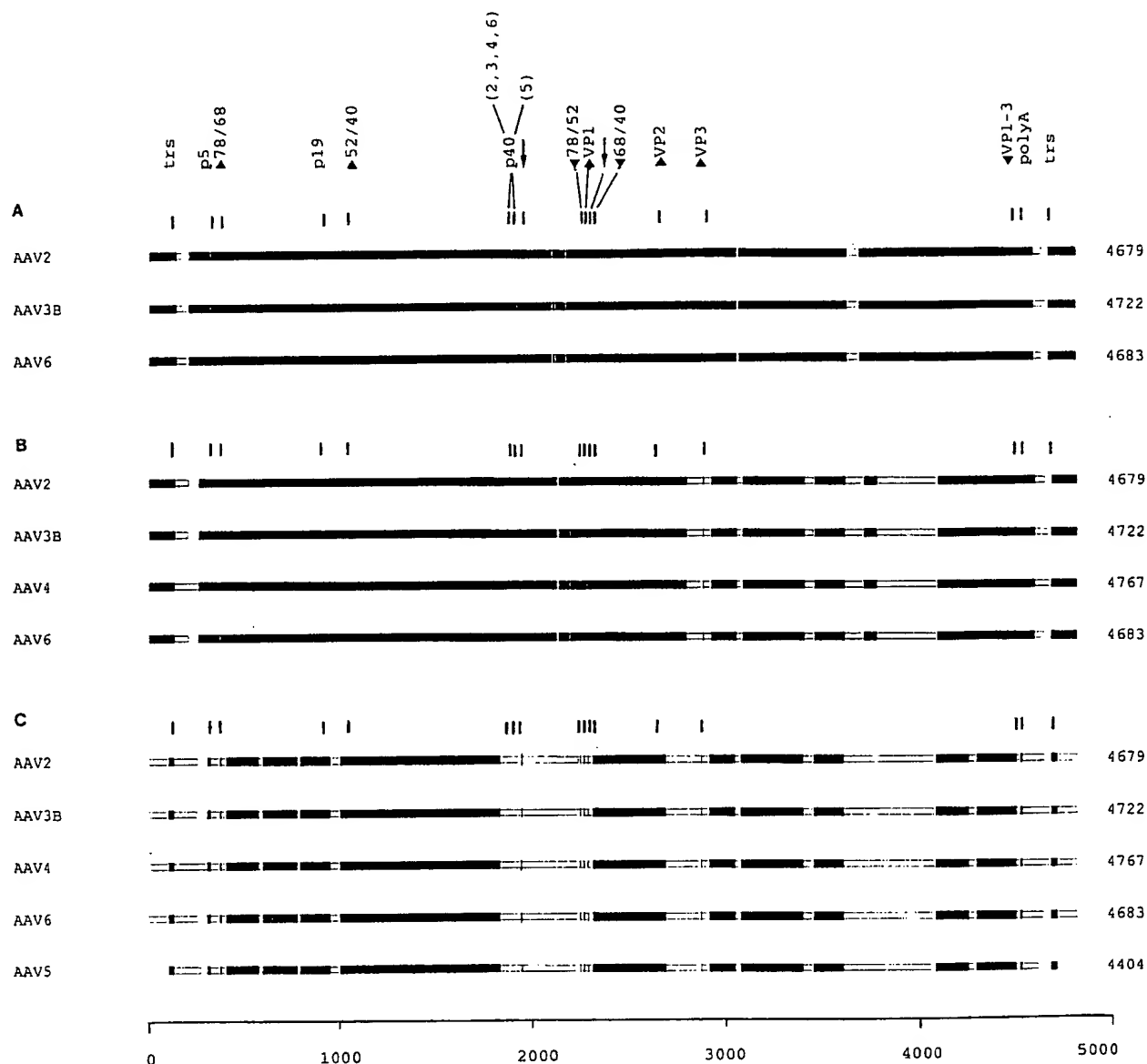
FIG. 3. Block alignment of AAV genomic sequences. The complete sequences of AAV2, AAV3B, AAV4, and AAV6 and the partial sequence of AAV5 were aligned with the MACAW program. Segment pair overlap with a minimum segment pair score of 60 was used as a search method to define nucleotide blocks of local similarity. The aligned blocks were linked into the alignment (solid bars). To determine the start and end positions for the alignment of AAV5, the trs (50) were selected by sequence analogy. The scale of alignment is 5,000 nt. Nucleotide positions on the individual genomes are in relation to this scale. The panels show the alignment of AAV2, AAV3B, and AAV6 (A); the alignment of AAV2, AAV3B, AAV4, and AAV6 (B); and the alignment of AAV2, AAV3B, AAV4, AAV6, and AAV5 (C). The positions of the conserved genetic elements are marked, including the p5, p19, and p40 promoters (position of TATA boxes); the intron splice sites (GT/AG [arrows]); the polyadenylation signal (polyA); the translation start (up arrowheads) and stop (down arrowheads) sites for the Rep (Rep 78, 68, 52, and 40) and capsid (VP1 to VP3) proteins; and the terminal resolution site and its inverted counterpart (both marked trs). Note that all of the elements in panel A were within blocks defined by the program and that in panel B only the start of VP3 was excluded, while in panel C several were no longer in regions that met the criteria required for block selection [p5, p40, start Rep 78 and 68, stop Rep 78 and 52, start VP1 and VP3, the splice signals, and the poly(A) site]. Such positions were marked manually. Note that the p40 sites in panel C are located in stretches of high dissimilarity and were not aligned.

The location of the TATA boxes, splice sites, and start and stop codons indicate a genomic organization similar to that of other sequenced AAVs (11, 17, 36, 41, 42, 51). Promoters are found around the expected map position for p5, p19, and p40, and there is only one single polyadenylation site at map position 4284 (AATAAA) of the partial AAV5 sequence. Translation start and stop codons for all putative AAV5 nonstructural proteins (Rep) and capsid proteins (VPs), including the unusual ACG start codon for VP2 and one common stop codon shared by all three known capsid proteins, are at the expected locations.

Although up to now we were not able to determine the complete terminal sequences of AAV5, the approximate length of the genome could be deduced. On the basis of the position of the trs (Fig. 2), about 100 additional nucleotides are expected at each end of the DNA molecule. Thus, some 150 to 200 nt presumably would complete the 4,404 nt of the partial sequence. The resulting total of 4,550 to 4,600 nt is in agree-
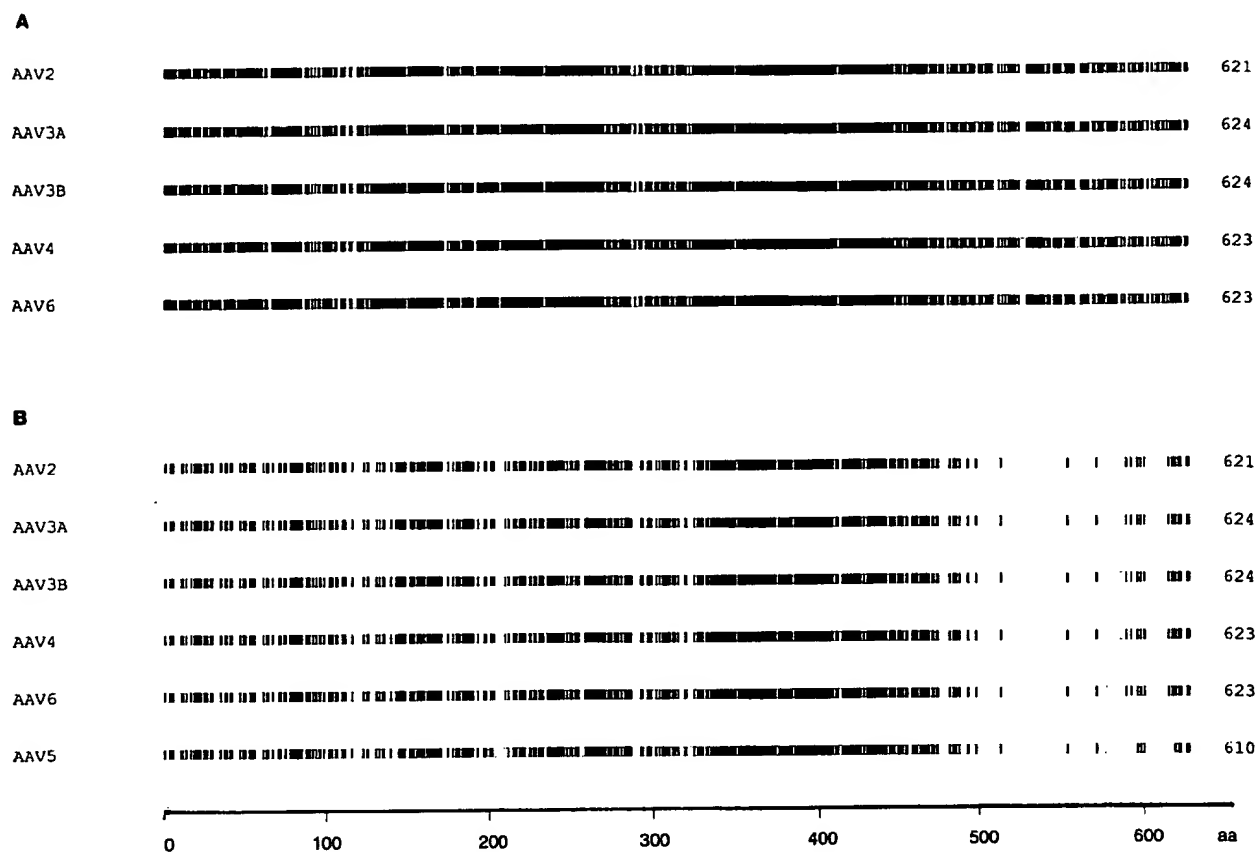
A

B

FIG. 4. Comparison of the left-sided ORFs of AAV. The left-sided ORFs of AAV were aligned with the MACAW program. Segment pair overlap with a minimum segment pair score of 60 was used as a search method to define amino acid (aa) blocks of local similarity. The aligned blocks were linked and identical amino acid residues were shaded. Shown are alignments of AAV2, AAV3A, AAV3B, AAV4, and AAV6 (A) and AAV2, AAV3A, AAV3B, AAV4, AAV6, and AAV5 (B).

ment with the size 4.5 to 4.6 kb derived from agarose gels and the mapping of restriction fragments (5).

To obtain data on binding sites for transcription factors, we made use of the TRANSFAC database (59). As with other AAVs (8, 17, 36, 42, 51), several putative binding sequences, such as CCAAT, GGGCGG, and GGTGGT boxes were found for AAV5. Others, such as the cyclic AMP-responsive element (CRE; TGACGTCA [20]), were found in AAV5 but not in AAV2 DNA or were located at unusual map positions, e.g., the consensus sequence GTGACGT for the transcription factor EivF (18, 39). In all AAVs sequenced up to now this sequence was found upstream of the p5 region (8, 17, 36, 42, 51), but in the case of AAV5 it is at map position 1740 and thus is upstream of the p40 promoter. Other recognition sites, such as the sequence targets for YY1 (CGACATTTT or CTCCATT TT) near the p5 promoter of AAV2 or the p7 promoter of AAV4 (17, 49), were not found in AAV5 DNA.

Similarities among AAV genomes. To find regions of local similarity among the different sequences and to define blocks of aligned sequence segments, we made use of the MACAW alignment program (National Center for Biotechnology Information). The different multiple alignments are shown in schematic form in Fig. 3.

When the sequences of AAV2, AAV3B, and AAV6 are compared (score cutoff, 60), almost the entire genomes could be well aligned into blocks, with the exception of small regions between the terminal repeats and the transcribed part of the genomes, some heterogeneity in the splice region, and one

interval corresponding to AAV2 nt 3548 to 3610 (Fig. 3A). These results are in agreement with the 82% homology reported by Muramatsu et al. between AAV2 and AAV3A (36) and by Rutledge et al. among AAV2, AAV3B, and AAV6 (42).

AAV4 is more distantly related to the genomes of AAV2, AAV3B, and AAV6 (17, 42). Thus, the inclusion of AAV4 in the alignment results in the extension and addition of regions of low similarity, especially in the right half of the genomes (Fig. 3B).

Since we did not succeed in reading the entire sequence of the AAV5 genome, we used the terminal resolution site and the respective inverted sequence (50) (Fig. 2B) as reference positions for aligning the partial AAV5 sequence to the published DNA sequences of AAV2, AAV3B, AAV4, and AAV6 (11, 17, 36, 41, 42, 51). Alignment of all five sequences resulted in a further reduction of regions with similarities above the cutoff (Fig. 3C). The scattered blocks of sequence similarity around the splice region in the left half of the genome (Fig. 3A and B) are replaced by a large interval spanning more than 400 nt (nt 1778 to 2249 on the genome of AAV2 corresponding to nt 1684 to 2130 on the incomplete AAV5 genome). This region is also covered by a BamHI restriction fragment (5) used to design AAV5-specific primers (54).

We also used pairwise alignments to find out whether AAV5 may be more closely related to one or more of the other members of the AAV family. But all homologies of the four pairs were between 54 and 56%, and the obtained patterns (data not shown) were comparable to the picture shown for the
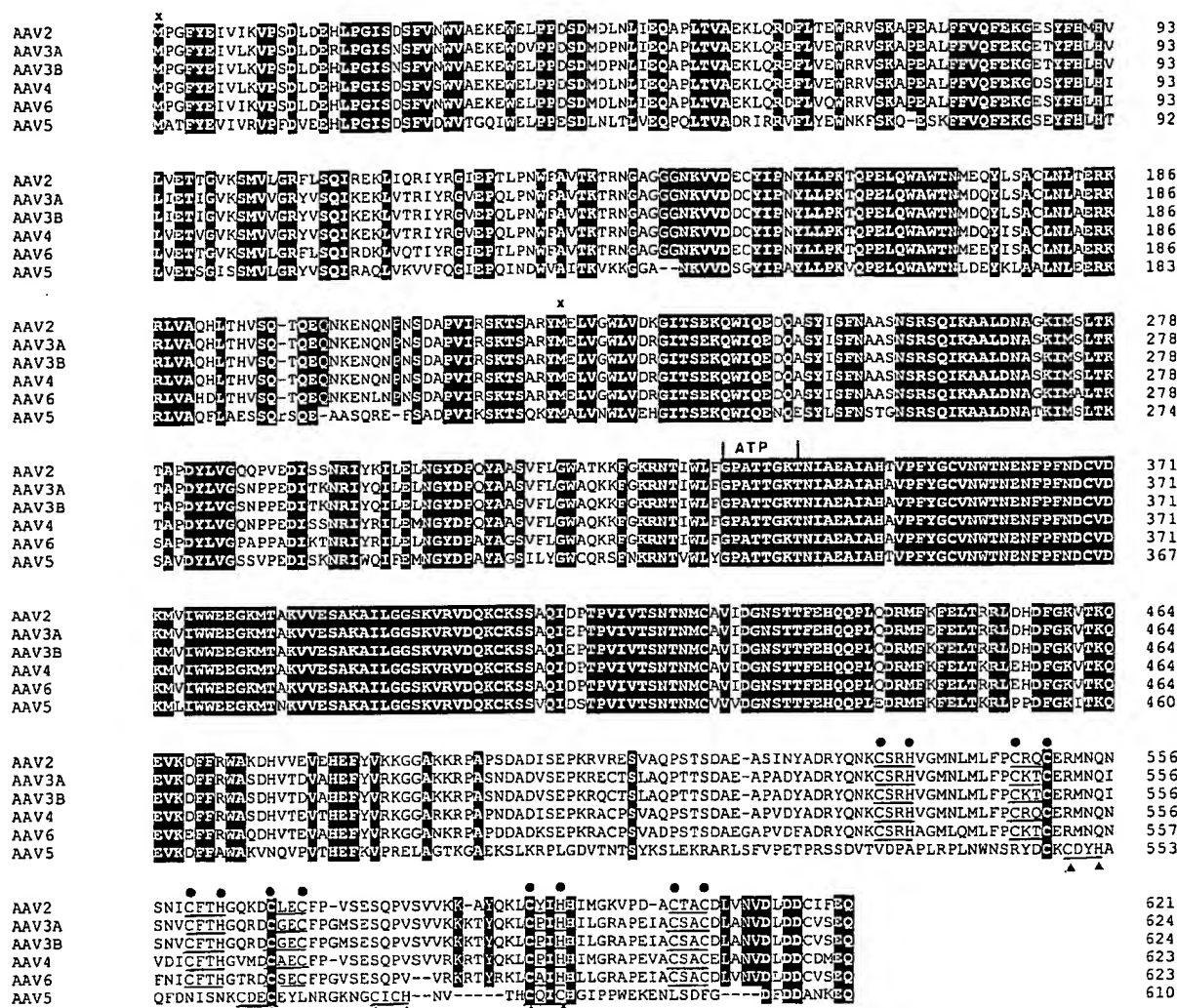
```
                                                                                      x
AAV2    MPGFYEIVIKVPSDLDEHLPGISDSFVNWVAEKEWELPPDSDMDLNLIEQAPLTVAEKLQRDFLTEWRRVSKAPEALFFVQFEKGESYFHLHV   93
AAV3A   MPGFYEIVLKVPSDLDERLPGISNSFVNWVAEKEWELPPDSDMDPNLIEQAPLTVAEKLQREDVPPDSDMDPNLIEQAPLTVAEKLQRETYFHLHV   93
AAV3B   MPGFYEIVLKVPSDLDERLPGISNSFVNWVAEKEWELPPDSDMDPNLIEQAPLTVAEKLQREPLVEWRRVSKAPEALFFVQFEKGETYFHLHV   93
AAV4    MPGFYEIVLKVPSDLDEHLPGISDSFVSWVAEKEWELPPDSDMDLNLIEQAPLTVAEKLQREFLVEWRRVSKAPEALPFVQFEKGDSYFBLHI   93
AAV6    MPGFYEIVIKVPSDLDEHLPGISDSFVNWVAEKEWELPPDSDMDLNLIEQAPLTVAEKLQRDFLVQWRRVSKAPEALFFVQFEKGESYFBLHI   93
AAV5    MATFYEVIVRVPFDVEEHLPGISDSFVDTVTGQIWELPPESDLNLTLVEQPQLTVADRIRREVFLYEWNKFSRQ-ESKFFVQFEKGSEYFHLHT   92


AAV2    DVETTGVKSMVLGRFLSQIREKLIQRIYRGIEPTLPNWFAVTKTRNGAGGGNKVVDECYIPNYLLPKTQPELQWAWTNMEQYLSACLNLTERR  186
AAV3A   DIETTIGVKSMVVGRYVSQIKEKLVTRIYRGVEPQLPNWFAVTKTRNGAGGGNKVVDDCYIPNYLLPKTQPELQWAWTNMDQYLSACLNLAERR  186
AAV3B   DIETTIGVKSMVVGRYVSQIKEKLVTRIYRGVEPQLPNWFAVTKTRNGAGGGNKVVDDCYIPNYLLPKTQPELQWAWTNMDQYISACLNLAERR  186
AAV4    DVETVGVKSMVVGRYVSQIKEKLVTRIYRGVEPQLPNWFAVTKTRNGAGGGNKVVDECYIPNYLLPKTQPELQWAWTNMDQYLSACLNLAERR  186
AAV6    DVETTGVKSMVLGRFLSQIRDKLVQTIYRGIEPTLPNWFAVTKTRNGAGGGNKVVDECYIPNYLLPKTQPELQWAWTNMEEYISACLNLAERR  186
AAV5    DVETSGISSMVLGRYVSQIRAQLVKVVFQGIEPQINDWVAITKVKKGA--NKVVDSGYIPAYLLPKVQPELQWAWTNLDEYKLAALNLEERR  183


                                                                x
AAV2    RLVAQHLTHVSQ-TQEQNKENQNENSDAPVIRSKTSARYMELVGWLVDKGITSEKQWIQEDQASYISFNAASNSRSQIKAALDNAGKIMSLTR  278
AAV3A   RLVAQHLTHVSQ-TQEQNKENQNENSDAPVIRSKTSARYMELVGWLVDRGITSEKQWIQEDQASYISFNAASNSRSQIKAALDNASKIMSLTR  278
AAV3B   RLVAQHLTHVSQ-TQEQNKENQNENSDAPVIRSKTSARYMELVGWLVDRGITSEKQWIQEDQASYISFNAASNSRSQIKAALDNASKIMSLTR  278
AAV4    RLVAQHLTHVSQ-TQEQNKENQNENSDAPVIRSKTSARYMELVGWLVDRGITSEKQWIQEDQASYISFNAASNSRSQIKAALDNASKIMSLTR  278
AAV6    RLVAHDLTHVSQ-TQEQNKENLNENSDAPVIRSKTSARYMELVGWLVDRGITSEKQWIQEDQASYISFNAASNSRSQIKAALDNAGKIMALTR  278
AAV5    RLVAQFLAESSQISQE-AASQRE-FSADPVIKSKTSQKYMALVNWLVEHGITSEKQWIQENQESYLSFNSTGNSRSQIKAALDNATKIMSLTK  274


                                                                              | ATP |
AAV2    TAPDYLVGQQPVEDISSNRIYKILELNGYDPQYAASVFLGWATKKGGKRNTIWLFGPATTGKTNIAEAIAHTVPFYGCVNWTNENFPFNDCVD  371
AAV3A   TAPDYLVGSNPPEDITKNRIYQILELNGYDPQYAASVFLGWAQKKGGKRNTIWLFGPATTGKTNIAEAIABAVPFYGCVNWTNENFPFNDCVD  371
AAV3B   TAPDYLVGSNPPEDITKNRIYQILELNGYDPQYAASVFLGWAQKKGGKRNTIWLFGPATTGKTNIAEAIABAVPFYGCVNWTNENFPFNDCVD  371
AAV4    TAPDYLVGQNPPEDISSNRIYRILEMNGYDPQYAASVFLGWAQKKGGKRNTIWLFGPATTGKTNIAEAIABAVPFYGCVNWTNENFPFNDCVD  371
AAV6    SAPDYLVGPAPPADIKTNRIYRILELNGYDPAYAGSVFLGWAQKRFGGKRNTIWLFGPATTGKTNIAEAIABAVPFYGCVNWTNENFPFNDCVD  371
AAV5    SAVDYLVGSSVPEDISKNRIYWQIREMNGYDPAYAGSILYGWCQRSENKRNTVWLYGPATTGKTNIAEAIAHTVPFYGCVNWTNENFPFNDCVD  367


AAV2    KMVIWWEEGKMTAKVVESAKAILGGSKVRVDQKCKSSAQIDPTPVIVTSNTNMCAVIDGNSTTFEHQQPLQDRMFKFELTRRLDHDFGKVTKQ  464
AAV3A   KMVIWWEEGKMTAKVVESAKAILGGSKVRVDQKCKSSAQIEPTPVIVTSNTNMCAVIDGNSTTFEHQQPLQDRMFEFELTRRLDHDFGKVTKQ  464
AAV3B   KMVIWWEEGKMTAKVVESAKAILGGSKVRVDQKCKSSAQIDPTPVIVTSNTNMCAVIDGNSTTFEHQQPLQDRMFKFELTRRLEHDFGKVTKQ  464
AAV4    KMVIWWEEGKMTAKVVESAKAILGGSKVRVDQKCKSSAQIDPTPVIVTSNTNMCAVIDGNSTTFEHQQPLQDRMFKFELTRRLEHDFGKVTKQ  464
AAV6    KMVIWWEEGKMTAKVVESAKAILGGSKVRVDQKCKSSAQIDPTPVIVTSNTNMCAVIDGNSTTFEHQQPLQDRMFKFELTRRLEHDFGKVTKQ  464
AAV5    KMLIWWEEGKMTNKVVESAKAILGGSKVRVDQKCKSSVQIDSTPVIVTSNTNMCVVVDGNSTTFEHQQPLEDRMFKFELTKRLPPDFGKITKQ  460


                                                              • •            • •
AAV2    EVKDFFRWAKDHVVEVEHEFYVKKGGAKKRPAPSDADISEPKRVRESVAQPSTSDAE-ASINYADRYQNKCSRHVGMNLMLFPCRQGERMNQN  556
AAV3A   EVKDFFRWASDHVTDVAHEFYVRKGGAKKRPASNDADVSEPKRECTSLAQPTTSDAE-APADYADRYQNKCSRHVGMNLMLFPCKTGERMNQI  556
AAV3B   EVKDFFRWASDHVTDVAHEFYVRKGGAKKRPASNDADVSEPKRQCTSLAQPTTSDAE-APADYADRYQNKCSRHVGMNLMLFPCKTGERMNQI  556
AAV4    EVKDFFRWASDHVTEVAHEFYVRKGGARKRPAPNDADISEPKRACPSVAPVDYADRYQNKCSRHVGMNLMLFPCRQGERMNQN  556
AAV6    EVKEFFRWAQDHVTEVAHEFYVRKGGANKRPAPDDADKSEPKRACPSVADPSTSDAEGAPVDFADRYQNKCSRHAGMLQMLFPCKTGERMNQN  557
AAV5    EVKDFFAWAKVNQVPVTHEFKVPRELAGTKGAEKSLKRPLGDVTNTSYKSLEKRARLSFVPETPRSSDVTVDPAPLRPLNWNSRYDGKCDYHA  553
                                                                                         ▲  ▲


        • •    • •                                     • •             • •
AAV2    SNICFTHGQKDGLECFP-VSESQPVSVVKK-AHQKLCYIHHIMGKVPD-ACTACDLVNVDLDDCIFEQ  621
AAV3A   SNVCFTHGQRDCGECFPGMSESQPVSVVKKKTYQKLCPIHHILGRAPEIACSACDLANVDLDDCVSEQ  624
AAV3B   SNVCFTHGQRDCGECFPGMSESQPVSVVKKKTYQKLCPIHHILGRAPEIACSACDLANVDLDDCVSEQ  624
AAV4    VDICFTHGVMDCAECFP-VSESQPVSVVRKRTYQKLCPIHHIMGRAPEVACSACELANVDLDDCDMEQ  623
AAV6    FNICFTHGTRDCSECFPGVSESQPV--VRKRTRKLCAIHHLLGRAPEIACSACDLVNVDLDDCVSEQ  623
AAV5    QFDNISNKCDECEYLNRGKNGCICH--NV-----THCQICEGIPPWEKENLSDFG----DFDDANKEQ  610
        ▲     ▲        ▲    ▲                            ▲ ▲
```

FIG. 5. Putative amino acid sequences of the Rep ORFs. The Rep ORFs of AAV2, AAV3A, AAV3B, AAV4, AAV6, and AAV5 were compared by using the MACAW alignment program. Identical amino acids are shown with white letters on a black background. Dashes indicate gaps in the sequence added by the alignment program. The putative ATP-binding site (ATP) and Zn fingers (three pairs of black dots above the sequences for all AAVs except AAV5, and two pairs of black triangles below the sequences for AAV5) are marked. ×, start of the two unspliced Rep proteins.

overall alignment in Fig. 3C. Thus, AAV5 appears to be more distant from the rather closely related group of AAV2, AAV3, AAV4, and AAV6. This finding is in agreement with the hybridization data reported in 1984 (5).

Nonstructural (Rep) protein coding region. By analogy with other AAVs, the AAV5 left-sided ORF encodes the putative unspliced nonstructural Rep proteins. The complete ORF (nt 239 to 2068) encodes a protein of 610 amino acids, and the additional ATG start codon (nt 899) would give rise to a smaller unspliced protein of 390 amino acids (nt 899 to 2068). Thus, both putative unspliced AAV5 Rep proteins are slightly smaller than those reported for the other AAV serotypes (8, 17, 36, 37, 42).

When the aligned left-sided ORFs of AAV2, AAV3A, AAV3B, AAV4, and AAV6 were compared at the amino acid level, the overall identity was 83.4% (data not shown; see also Fig. 4A). However, the percentage of identical amino acid residues was reduced to 54.4% when AAV5 was included in the alignment (see also Fig. 4B and 5). When Fig. 4A is compared with Fig. 4B, it becomes evident that the amount of

identical amino acids was reduced throughout the sequence, but it was most striking for the sequences from amino acid 483 of AAV5 (corresponding to residue 487 of AAV2) to the C-terminal amino acid 610 (621 of AAV2). Whereas without AAV5 the identity in this part of the aligned sequences could still be attributed to 91 of 135 (67.4%) of the amino acids, the value was only 13 identical residues of 128 (10.1%) when it was included (see Fig. 5).

The differences between the AAV5 noncapsid proteins and the respective proteins of the other AAV serotypes may affect the secondary structure and functional domains. While the putative ATP-binding site (47) is well conserved in all AAVs, one of the three putative Zn fingers at the carboxyl terminus (for a review, see reference 10) is missing in the AAV5 Rep proteins (Fig. 5). The distances between the three Zn-binding motifs (CXXH/CXXC) in AAV2 are 9, 4, and 9 amino acids. The respective two motifs in AAV5 have distances of 9 and 4 amino acids.

Capsid protein coding region. The right-sided ORF (nt 2087 to 4258) encodes a capsid protein (VP1) of 724 amino acids,
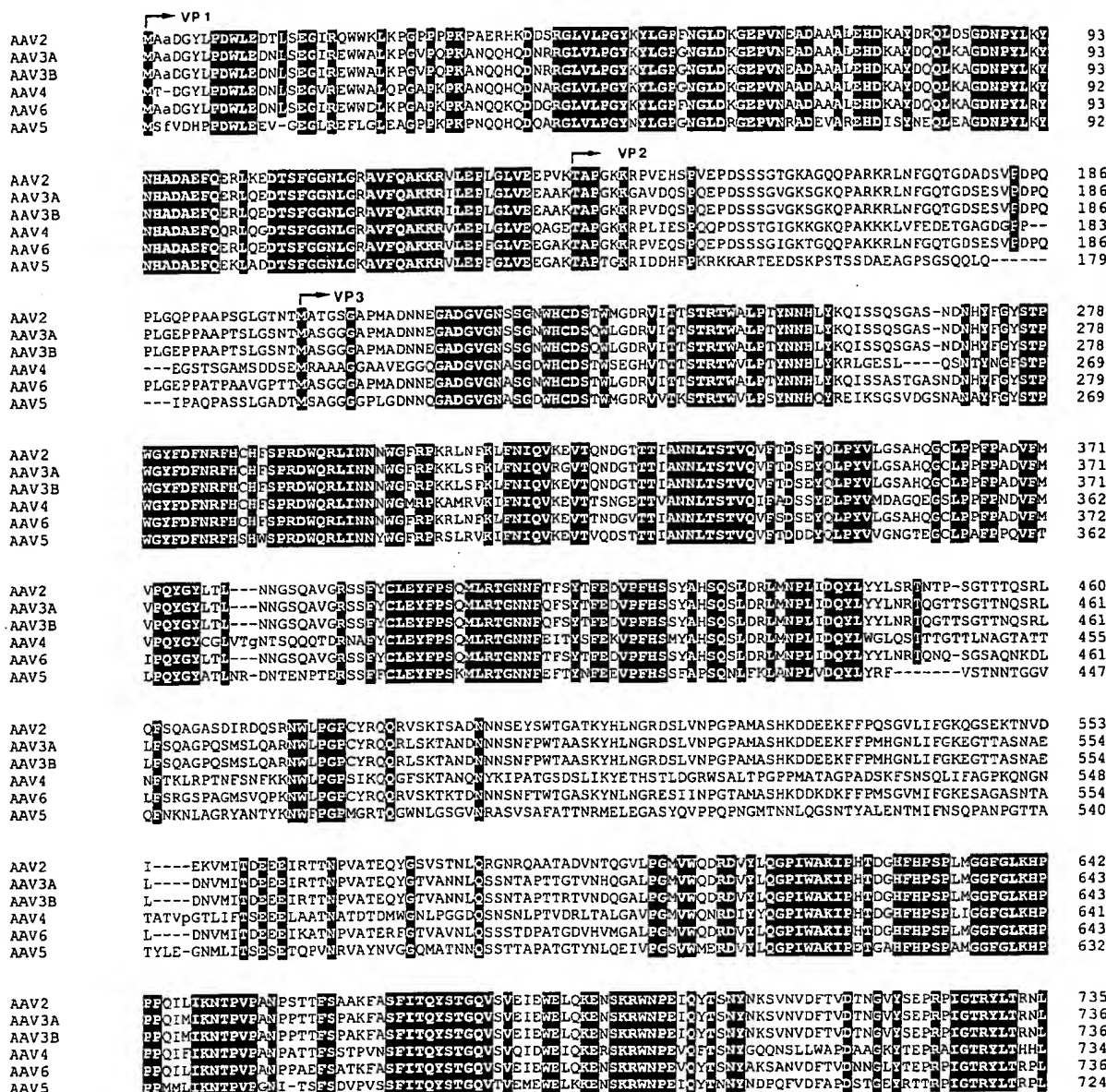
FIG. 6. Alignment of the right-sided cap ORFs of AAV. The right-sided ORFs of AAV were aligned with the MACAW program. Segment pair overlap with a minimum segment pair score of 60 was used as the search method to define amino acid blocks of local similarity. Identical amino acid residues are shown with white letters on a black background. VP1, VP2, and VP3 indicate the beginnings of the respective capsid proteins.

with a predicted molecular mass of 80 kDa. The VP2 capsid (nt 2495 to 4258), which by analogy with other AAVs is presumably initiated at an ACG triplet, and the VP3 capsid (nt 2663 to 4258) would comprise 588 and 532 amino acids, respectively, with corresponding molecular masses of 65 and 59 kDa. Thus, all of the molecular masses predicted for the AAV5 capsid proteins are smaller than those reported for other AAVs (8, 17, 36, 37, 42).

In a group alignment of the amino acid sequences of the capsid proteins of AAV2, AAV3, and AAV6, very high identities of more than 80% were found (Table 1). The similarity was considerably decreased to 59.7% when AAV4 was included in the alignment (Table 1). The pairwise alignments yielded identities between 61.5% for AAV2 versus AAV4 and 64.8% for AAV3B versus AAV4 (Table 1). Blocks of very high sim-

ilarity were obtained from amino acids 1 to 173 (AAV2) and amino acids 599 to 735 upon both pairwise and overall alignments of the respective sequences.

The results of pairwise alignments of capsid ORFs of AAV2, AAV3A, AAV3B, AAV4, and AAV6 to AAV5 showed identities between 51.4% for AAV4 versus AAV5 and 58.8% for AAV6 versus AAV5 (Table 1), and the alignment of all six capsid ORFs resulted in a reduction of overall amino acid identity to 44.9% (Fig. 6; Table 1). This is in correlation to the increased heterogeneity seen upon alignment of the respective nucleotide sequences (Fig. 3C).

As can be seen from the results listed in Table 1, including either AAV4 or AAV5 in the group alignments strongly reduced the percentage of identities between the capsid ORFs of the AAV serotypes 2, 3A, 3B, and 6 (from 80.1 to 59.7 and

TABLE 1. Percentage of identical amino acids in the right-sided ("cap") ORF derived from AAV alignments[a]

AAV serotype alignments	% Identical amino acids		
	Pairwise		Group
	Aligned to AAV4	Aligned to AAV5	
Pairwise			
AAV2	61.5	58.3	
AAV3A	63.9	58.4	
AAV3B	64.8	58.7	
AAV4	100	51.4	
AAV6	64.6	58.8	
Group			
AAV2, AAV3A, AAV3B, and AAV6			80.1
AAV2, AAV3A, AAV3B, and AAV6 versus AAV4			59.7
AAV2, AAV3A, AAV3B, and AAV6 versus AAV5			53.2
AAV2, AAV3A, AAV3B, and AAV6 versus AAV4 and AAV5			44.9

[a] The right-sided AAV ORFs were aligned either pairwise or in the indicated combinations with the MACAW program, and the identical amino acids were determined.

53.2%, respectively). However, the pairwise alignment of the capsid ORFs of AAV4 and AAV5 yields already a lower degree of similarity (51.4; Table 1). Thus, the capsid ORFs of both AAV4 and AAV5 differ from the other serotypes as well as from each other. Therefore, the percentage of identical amino acids decreases further (44.9%) when both AAV4 and AAV5 are included in the group alignment (Table 1).

DISCUSSION

We have determined the sequence of 4,404 nt (about 95% of the estimated size of 4.5 to 4.6 kb) of the AAV5 genome, but up to now we were not able to resolve the complete sequence of the terminal repeats. Using the trs (50), the respective inverted sequence, and the TATAA box in the p5 promoter as reference points, we have made alignments of the AAV5 sequence to the published sequences of AAV2, AAV3B, AAV4, and AAV6 (17, 36, 41, 42, 51). The overall organization within the AAV5 genome is similar to that of other AAVs with three putative promoters, one single polyadenylation site, and two large ORFs.

As reported by others, the identities on the nucleotide level between AAV2, AAV3, and AAV6 exceeded 82% and were still as high as 75% when AAV4 was included (17, 36, 42). In contrast to the good sequence conservation between the other AAV serotypes, AAV5 is clearly more distantly related, and the overall identity is reduced to about 56% when its nucleotide sequence is included in the alignment. Thus, the distance between AAV5 and the other AAV serotypes is in the same range as that reported for the other AAVs compared to their closest relatives among autonomous parvoviruses, namely, the goose and duck parvoviruses (17, 36, 62).

The differences in the nucleotide sequence near the p5 and the p40 promoters appear to be connected to differences in the adenovirus transcription factor recognition sites. Thus, the EivF element involved in E1A-responsive transactivation of the adenovirus E4 promoter (18, 39) is shifted from the common region upstream of the p5 promoter to a position upstream of the p40 promoter in AAV5, and the YY1 sites (49)

are not found in the AAV5 genome. If and how these consensus sequences and other sequences (e.g., the CRE element [20]) are involved in adenovirus transactivation remains to be determined.

Within the coding regions, stretches of dissimilarity were most prominent at AAV5 nt 1684 to 2130 (AAV2 nt 1778 to 2249), 2504 to 2692 (AAV2 nt 2623 to 2839), and 3393 to 3848 (AAV2 nt 3529 to 3993). The last two affect all three capsid proteins. Thus, the capsid proteins of AAV5 differ significantly from the respective proteins of the other serotypes, and the overall percentage of identical amino acids is less than 45% (Table 1). Since the respective differences comprise regions which are supposed to be exposed at the virus surface (12), tissue tropism, cellular receptor(s), and resistance towards AAV neutralizing antibodies might be different from those of other AAVs. Thus, the host range of AAV5 may be distinct. This should be of interest for the construction of AAV gene transduction vectors and for their application in gene therapy (31, 37, 48).

The region of nucleotide sequence dissimilarity at AAV5 map positions 1684 to 2130 (AAV2 nt 1778 to 2249) reflects differences in the Rep proteins. This certainly represents the most striking finding, since the sequence of Rep proteins was thought to be highly conserved because of the essential functions that these proteins have in the AAV life cycle. Whether the altered sequence of the AAV5 Rep proteins and the lack of a third Zn finger motif leads to structurally and functionally different proteins remains to be elucidated. It is noteworthy that the alignment of each of the AAV sequences to the respective sequences of goose parvovirus (4, 62) showed a high sequence conservation around the Rep ATP-binding site (47). This result (4) is in agreement with the high degree of sequence similarity seen at the respective positions when the Rep ORFs of all AAVs are aligned (Fig. 5). Thus, while the C-terminal sequences and the Zn fingers may be more flexible, the high degree of conservation of the ATP-binding site suggests its unique structure and points to its central role in the viral life cycle.

REFERENCES

1. Atchison, R. W., B. C. Casto, and W. M. Hammon. 1965. Adenovirus-associated defective virus particles. Science 194:754–756.
2. Bachmann, P. A., M. D. Hoggan, J. L. Melnick, H. G. Pereira, and C. Vago. 1975. Parvoviridae. Intervirology 5:83–92.
3. Bantel-Schaal, U. 1995. Growth properties of a human melanoma cell line are altered by adeno-associated parvovirus type 2. Int. J. Cancer 60:269–274.
4. Bantel-Schaal, U., and H. Delius. 1998. Unpublished results.
5. Bantel-Schaal, U., and H. zur Hausen. 1984. Characterization of the DNA of a defective human parvovirus isolated from a genital site. Virology 134:52–63.
6. Beaton, A., P. Palumbo, and K. I. Berns. 1989. Expression from the adeno-associated virus p5 and p19 promoters is negatively regulated in trans by the rep protein. J. Virol. 63:4450–4454.
7. Berns, K. I. 1990. Parvovirus replication. Microbiol. Rev. 54:316–329.
8. Berns, K. I. 1995. Parvoviridae: the viruses and their replication, p. 1017–1041. In B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), Fundamental virology, 3rd ed. Lippincott-Raven Publishers, Philadelphia, Pa.
9. Berns, K. I., and M. A. Labow. 1987. Parvovirus gene regulation. J. Gen. Virol. 68:601–614.
10. Carter, B. J., J. P. Trempe, and E. Mendelson. 1990. Adeno-associated virus gene expression and regulation, p. 227–254. In P. Tijsen (ed), Handbook of parvoviruses, vol. 1. CRC Press, Boca Raton, Fla.
11. Cassinotti, P., M. Weitz, and J. D. Tratschin. 1988. Organization of the adeno-associated virus (AAV) capsid gene: mapping of a minor spliced mRNA coding for virus capsid protein 1. Virology 167:176–184.
12. Chapmann, M. S., and M. G. Rossmann. 1993. Structure, sequence, and function correlations among parvoviruses. Virology 194:491–508.
13. Chejanovsky, N., and B. J. Carter. 1989. Mutagenesis of an AUG codon in the adeno-associated virus rep gene: effects on viral DNA replication. Virology 171:120–128.
14. Chejanovsky, N., and B. J. Carter. 1990. Mutation of a consensus purine nucleotide binding site in the adeno-associated virus rep gene generates a dominant-negative phenotype for DNA replication. J. Virol. 51:329–339.

15. Chiorini, J. A., S. M. Wiener, R. M. Kotin, R. A. Owens, S. R. M. Kyöstiö, and B. Safer. 1994. Sequence requirements for stable binding and function of Rep68 on the adeno-associated virus type 2 inverted terminal repeats. J. Virol. 68:7448–7457.

16. Chiorini, J. A., L. Yang, B. Safer, and R. M. Kotin. 1995. Determination of adeno-associated virus Rep68 and Rep78 binding sites by random sequence oligonucleotide selection. J. Virol. 69:7334–7338.

17. Chiorini, J. A., L. Yang, Y. Liu, B. Safer, and R. M. Kotin. 1997. Cloning of adeno-associated virus type 4 (AAV4) and generation of recombinant AAV4 particles. J. Virol. 71:6823–6833.

18. Cortes, P., L. Buckbinder, M. A. Leza, N. Rak, P. Hearing, A. Merino, and D. Reinberg. 1988. EivF, a factor required for transcription of the adenovirus EIV promoter, binds to an element involved in E1a-dependent activation and cAMP induction. Genes Dev. 2:975–990.

19. Georg-Fries, B., S. Biederlack, J. Wolf, and H. zur Hausen. 1984. Analysis of proteins, helper dependence, and seroepidemiology of a new human parvovirus. Virology 134:64–71.

20. Hardy, S., and T. Shenk. 1988. Adenoviral control regions activated by E1A and the cAMP response element bind to the same factor. Proc. Natl. Acad. Sci. USA 85:4171–4175.

21. Heilbronn, R., A. Bürkle, S. Stephan, and H. zur Hausen. 1990. The adeno-associated virus *rep* gene suppresses herpes simplex virus-induced DNA amplification. J. Virol. 64:3012–3018.

22. Hermonat, P. L. 1991. Inhibition of H-ras expression by the adeno-associated virus Rep78 transformation suppressor gene product. Cancer Res. 46:3373–3377.

23. Hermonat, P. L. 1994. Down regulation of the human *c-fos* and *c-myc* protooncogene promoters by adeno-associated virus Rep78. Cancer Lett. 81:129–136.

24. Hermonat, P. L. 1994. Adeno-associated virus inhibits human papillomavirus type 16: a viral interaction implicated in cervical cancer. Cancer Res. 54:2278–2281.

25. Hermonat, P. L., R. T. Plott, A. D. Santin, G. P. Parham, and J. T. Flick. 1997. Adeno-associated virus Rep78 inhibits oncogenic transformation of primary human keratinocytes by a human papillomavirus type 16-*ras* chimeric. Gynecol. Oncol. 66:487–494.

26. Hermonat, P. L., A. D. Santin, and R. B. Batchu. 1996. The adeno-associated virus Rep78 major regulatory/transformation suppressor protein binds Sp1 *in vitro* and evidence of a biological effect. Cancer Res. 56:5299–5304.

27. Hölscher, C., J. A. Kleinschmidt, and A. Bürkle. 1995. High-level expression of adeno-associated virus (AAV) Rep78 and Rep68 protein is sufficient for infectious-particle formation by a *rep*-negative AAV mutant. J. Virol. 69:6880–6885.

28. Hörer, M., S. Weger, K. Butz, F. Hoppe-Seyler, C. Geisen, and J. A. Kleinschmidt. 1995. Mutational analysis of adeno-associated virus Rep protein-mediated inhibition of heterologous and homologous promoters. J. Virol. 69:5485–5496.

29. Hoggan, M. D., N. R. Blacklow, and W. P. Rowe. 1966. Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics. Proc. Natl. Acad. Sci. USA 55:1467–1474.

30. Hunter, L. A., and R. J. Samulski. 1992. Colocalization of adeno-associated virus Rep and capsid proteins in the nuclei of infected cells. J. Virol. 66:317–324.

31. Kotin, R. M. 1994. Prospects for the use of adeno-associated virus as a vector for human gene therapy. Hum. Gene Ther. 5:793–801.

32. Kyöstiö, S. R., R. A. Owens, M. D. Weitzmann, B. Antoni, N. Chejanovsky, and B. J. Carter. 1994. Analysis of adeno-associated virus (AAV) wild-type and mutant Rep proteins for their abilities to negatively regulate AAV p5 and p19 mRNA levels. J. Virol. 68:2947–2957.

33. McCarty, D. M., D. J. Pereira, I. Zolotukhin, X. Zhou, J. H. Ryan, and N. Muzyczka. 1994. Identification of linear DNA sequences that specifically bind the adeno-associated virus Rep protein. J. Virol. 68:4988–4997.

34. Melnick, J. L., H. D. Mayor, K. O. Smith, and F. Rapp. 1965. Association of 20 millimicron particles with adenoviruses. J. Bacteriol. 90:271–274.

35. Mizukami, H., N. S. Young, and K. E. Brown. 1996. Adeno-associated virus type 2 binds to a 150-kilodalton cell membrane glycoprotein. Virology 217:124–130.

36. Muramatsu, S.-I., H. Mizukami, N. S. Young, and K. E. Brown. 1996. Nucleotide sequencing and generation of an infectious clone of adeno-associated virus 3. Virology 221:208–217.

37. Muzyczka, N. 1992. Use of adeno-associated virus as a general transduction vector for mammalian cells. Curr. Top. Microbiol. Immunol. 158:97–129.

38. Prasad, K.-M. R., and J. P. Trempe. 1995. The adeno-associated virus Rep78 protein is covalently linked to viral DNA in a preformed virion. Virology 214:360–370.

39. Raychaudhuri, P., R. Rooney, and J. R. Nevins. 1987. Identification of a E1A-inducible cellular factor that interacts with regulatory sequences within the adenovirus E4 promoter. EMBO J. 6:4073–4081.

40. Rommelaere, J., and J. J. Cornelis. 1991. Antineoplastic activities of parvoviruses. J. Virol. Methods 33:233–251.

41. Ruffing, M., H. Heid, and J. A. Kleinschmidt. 1994. Mutations in the carboxy terminus of adeno-associated virus 2 capsid proteins affect viral infectivity: lack of an RGD integrin-binding motif. J. Gen. Virol. 75:3385–3392.

42. Rutledge, E. A., C. L. Halbert, and D. W. Russell. 1998. Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2. J. Virol. 72:309–319.

43. Ryan, J. H., S. Zolotukhin, and N. Muzyczka. 1996. Sequence requirements for binding of Rep68 to the adeno-associated virus terminal repeats. J. Virol. 70:1542–1553.

44. Samulski, R. J. 1993. Adeno-associated virus: integration at a specific chromosomal locus. Curr. Opin. Genet. Dev. 3:74–80.

45. Samulski, R. J., X. Zhu, X. Xiao, J. D. Brook, D. E. Housman, N. Epstein, and L. A. Hunter. 1991. Targeted integration of adeno-associated virus (AAV) into human chromosome 19. EMBO J. 10:3941–3950.

46. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.

47. Saraste, M., P. R. Sibbald, and A. Wittinghofer. 1990. The P-loop: a common motif in ATP- and GTP-binding proteins. Trends Biochem. Sci. 15:430–434.

48. Shaughnessy, E., D. Lu, S. Chatterjee, and K. K. Wong. 1996. Parvoviral vectors for the gene therapy of cancer. Semin. Oncol. 23:159–171.

49. Shi, Y., E. Seto, L. S. Chang, and T. Shenk. 1991. Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. Cell 18:377–388.

50. Snyder, R. O., D. S. Im, T. Ni, X. Xiao, R. J. Samulski, and N. Muzyczka. 1993. Features of the adeno-associated virus origin involved in substrate recognition by the viral Rep protein. J. Virol. 67:6096–6104.

51. Srivastava, A., E. W. Lusby, and K. I. Berns. 1983. Nucleotide sequence and organization of the adeno-associated virus 2 genome. J. Virol. 45:555–564.

52. Su, P. F., and F. Y. Wu. 1996. Differential suppression of the tumorigenicity of HeLa and SiHa cells by adeno-associated virus. Br. J. Cancer 73:1533–1537.

53. Summerford, C., and R. J. Samulski. 1998. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. J. Virol. 72:1438–1445.

54. Tobiasch, E., T. Burguete, P. Klein-Bauernschmitt, R. Heilbronn, and J. R. Schlehofer. 1998. Discrimination between different types of human adeno-associated viruses in clinical samples by PCR. J. Virol. Methods 71:17–25.

55. Trempe, J. P., and B. J. Carter. 1988. Regulation of adeno-associated virus gene expression in 293 cells: control of mRNA abundance and translation. J. Virol. 62:68–74.

56. Trempe, J. P., E. Mendelson, and B. J. Carter. 1987. Characterization of adeno-associated virus rep proteins in human cells by antibodies raised against rep expressed in *Escherichia coli*. Virology 161:18–28.

57. Walz, C., and J. R. Schlehofer. 1992. Modification of some biological properties of HeLa cells containing adeno-associated-virus DNA integrated into chromosome 17. J. Virol. 66:2990–3002.

58. Weitzmann, M. D., S. R. M. Kyöstiö, R. M. Kotin, and R. A. Owens. 1994. Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. Proc. Natl. Acad. Sci. USA 91:5808–5812.

59. Wingender, E., A. E. Kel, O. V. Kel, H. Karas, T. Heinemeyer, P. Dietze, R. Knüppel, A. G. Romaschenko, and N. A. Kolchanov. 1997. TRANSFAC, TRRD and COMPEL: towards a federated database system on transcriptional regulation. Nucleic Acids Res. 25:265–268.

60. Wistuba, A., S. Weger, A. Kern, and J. A. Kleinschmidt. 1995. Intermediates of adeno-associated virus type 2 assembly: identification of soluble complexes containing Rep and Cap proteins. J. Virol. 69:5311–5319.

61. Wonderling, R. S., and R. A. Owens. 1996. The Rep68 protein of adeno-associated virus stimulates expression of the platelet-derived growth factor B c-*sis* proto-oncogene. J. Virol. 70:4783–4786.

62. Zádori, Z., R. Stefancsik, T. Rauch, and J. Kisary. 1995. Analysis of the complete nucleotide sequence of goose and muscovy duck parvoviruses indicates common ancestral origin with adeno-associated virus 2. Virology 212:562–573.